

Universidad del Valle de Guatemala

**Estudiantes:** Augusto Alonso, David Cuellar, Rudik Rompich

**Correos:** alo181085@uvg.edu.gt, cue18382@uvg.edu.gt, rom19857@uvg.edu.gt

**Carnés:** 181085, 18382, 19857

CC3066 - Data Science I - Catedrático: Luis Furlan

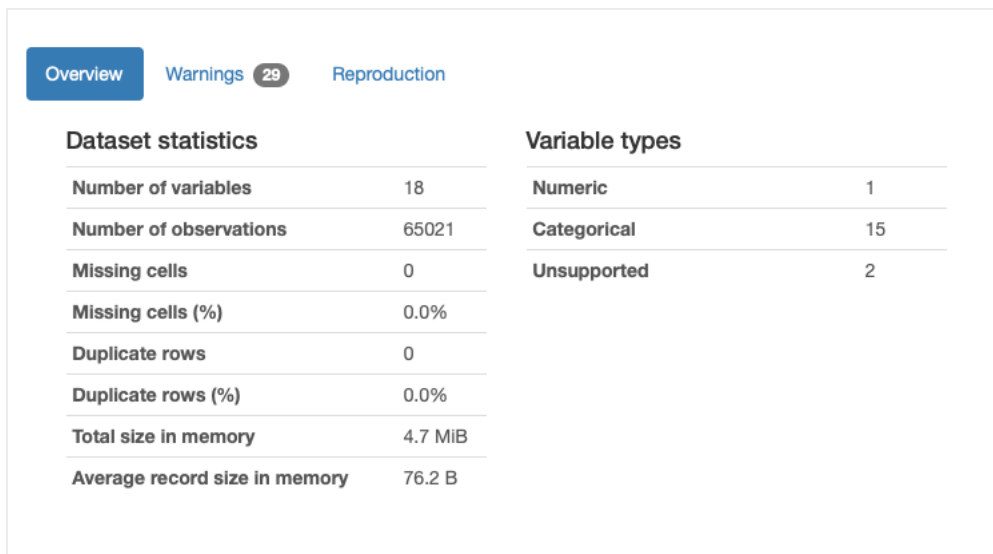
6 de agosto de 2021

---

## Code Book

### 1. Descripción general

El resumen de las estadísticas de la base de datos se puede observar en 1, la división de tipo de variables se eligió arbitrariamente; colocando a todas las variables como categóricas (exceptuando al código). La base de datos contiene 18 variables y tiene 65021 filas de todos los centros educativos.



The screenshot shows a software interface with three tabs: 'Overview' (active), 'Warnings' (with 29 warnings), and 'Reproduction'. Below the tabs are two tables. The first table, 'Dataset statistics', lists various metrics about the dataset. The second table, 'Variable types', shows the distribution of variable types.

Dataset statistics		Variable types	
Number of variables	18	Numeric	1
Number of observations	65021	Categorical	15
Missing cells	0	Unsupported	2
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	4.7 MiB		
Average record size in memory	76.2 B		

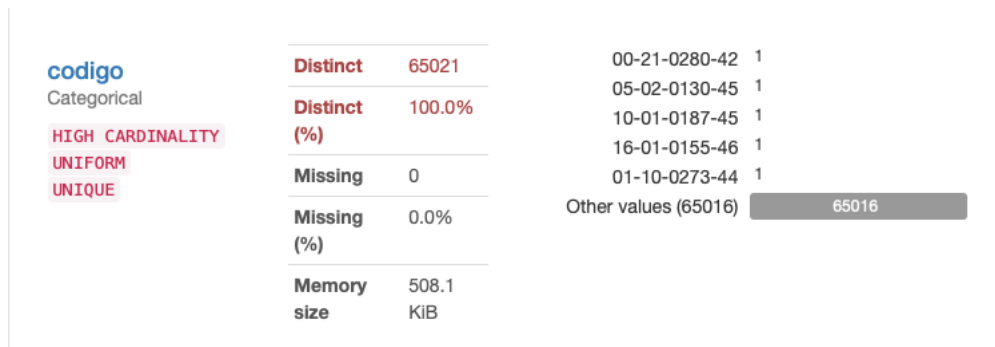
Figura 1: Resumen

## 2. Descripción de variables

**Variable 1.** (*codigo*) La variable código sirve para identificar cada establecimiento educativo por un índice único.

**Valores posibles.**

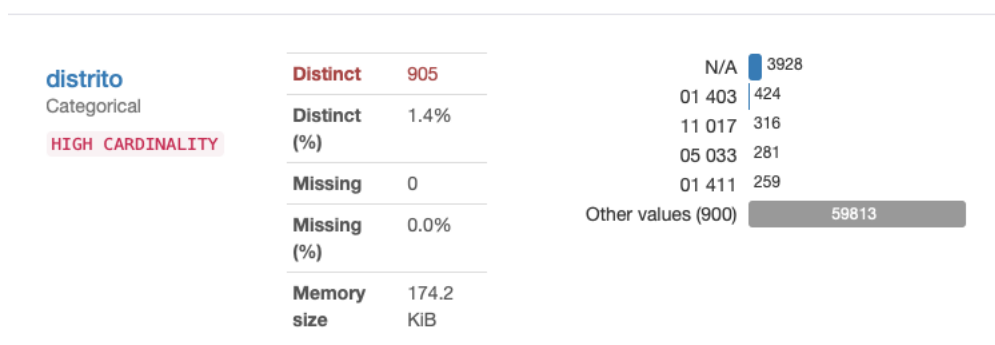
- 65021 índices distintos, representando a todos los centros educativos.



**Variable 2.** (*distrito*) Hace referencia al distrito donde se encuentra el centro educativo. Su utilidad se encuentra en que representa al distrito de supervisión al que se encuentra regido.

**Valores posibles.**

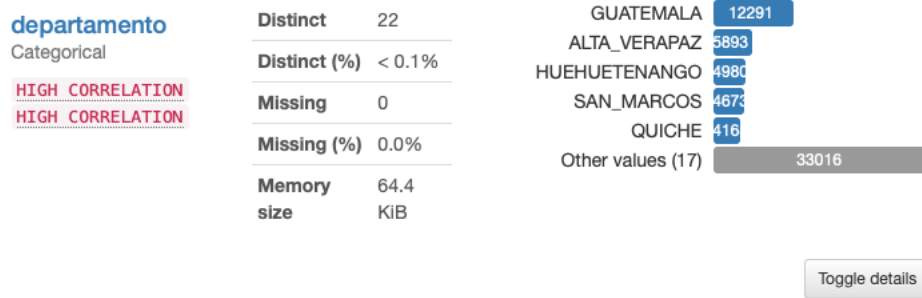
- 905 valores posibles.
- 3928 valores nulos.



**Variable 3.** (*departamento*) Es la variable que representa al departamento en donde se encuentra el centro educativo.

**Valores posibles.**

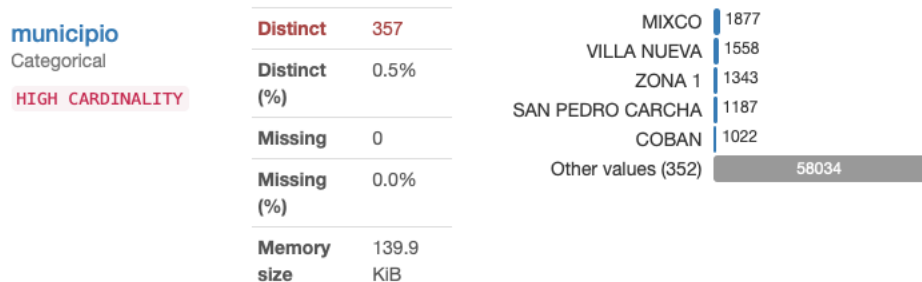
- Los 22 departamentos del país.



**Variable 4.** (*municipio*) Es la variable que representa al municipio en donde se encuentra el centro educativo.

**Valores posibles.**

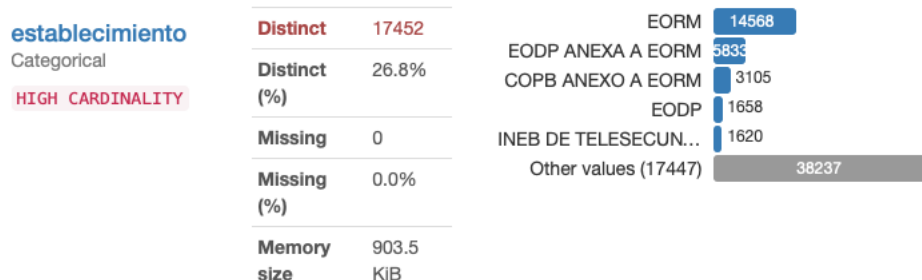
- Los 357 departamentos del país.



**Variable 5.** (*establecimiento*) Es una variable que hace referencia al nombre de los centros educativos y el tipo de centro educativo (*mixto, urbana, rural, por cooperativa, telesecundaria*). Los datos muestran que existen muchos centros educativos con el mismo nombre.

**Valores posibles.**

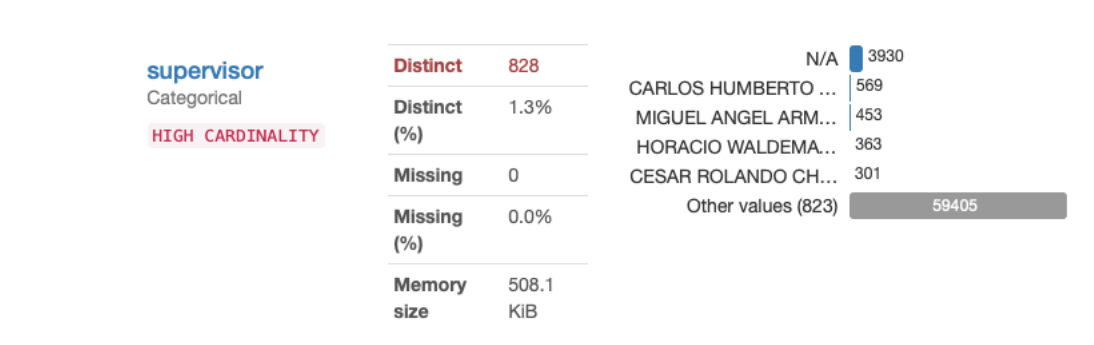
- 17452 centros educativos distintos.



**Variable 6.** (*supervisor*) Hace referencia al nombre de la persona que está encargada de la supervisión de los centros educativos.

**Valores posibles.**

- Existen 828 datos distintos.



**Variable 7.** (*director*) Es una variable que indica el nombre del director del centro educativo.

**Valores posibles.**

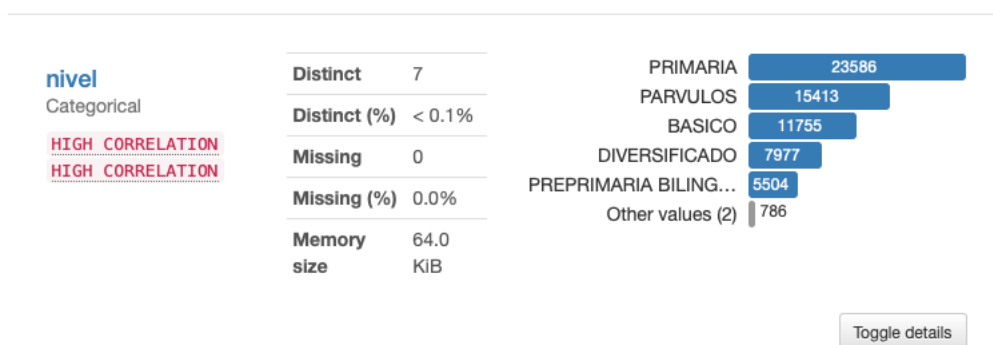
- Los valores posibles son 19087.
- La mayoría de valores son nulos.



**Variable 8.** (*nivel*) Es una variable que representa todos los niveles educativos que ofrecen los centros educativos.

**Valores posibles.**

- 7 valores distintos (desde primaria hasta diversificado).

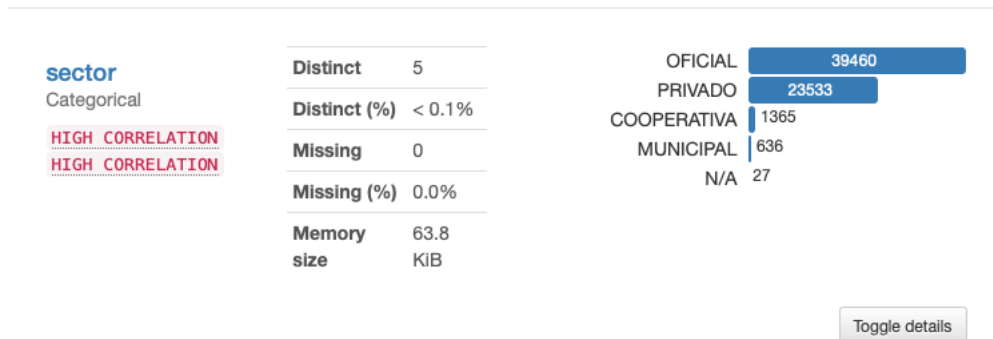


**Variable 9.** (*sector*) Es una variable que complementa a la variable de establecimiento ya que separa en una columna específica el tipo de centro de educativo que representa.

**Valores posibles.**

5 valores distintos.

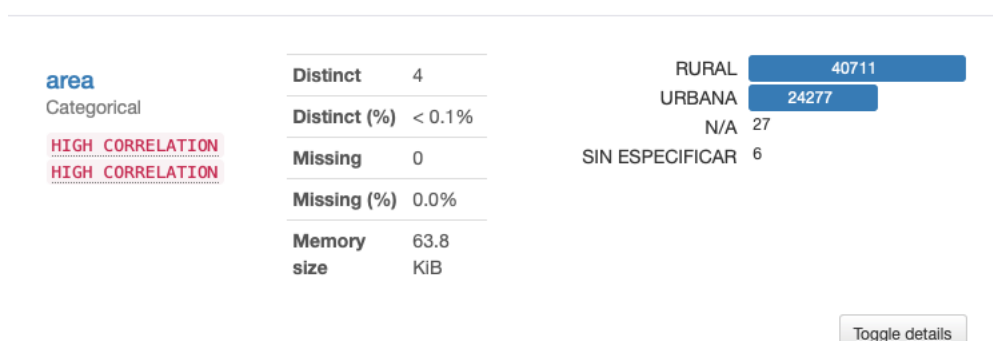
■



**Variable 10.** (*area*) Es una variable que complementa a la variable de establecimiento ya que separa en una columna específica la locación de centro de educativo.

**Valores posibles.**

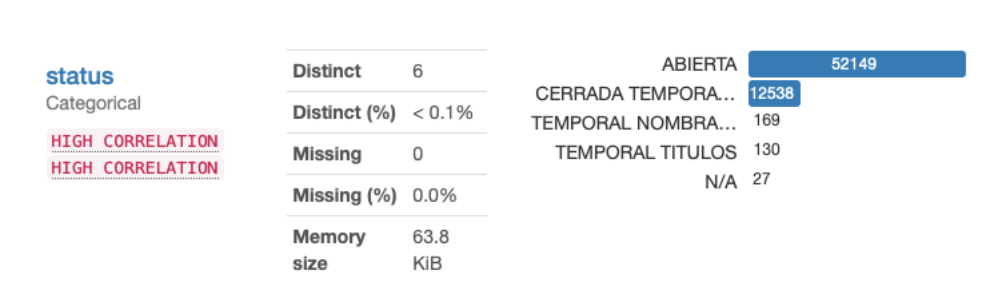
■ 4 valores distintos.



**Variable 11.** (*status*) Es una variable que representa el status actual del centro educativo, en caso de que se encuentre abierta, cerrado o cualquier otro pormenor.

**Valores posibles.**

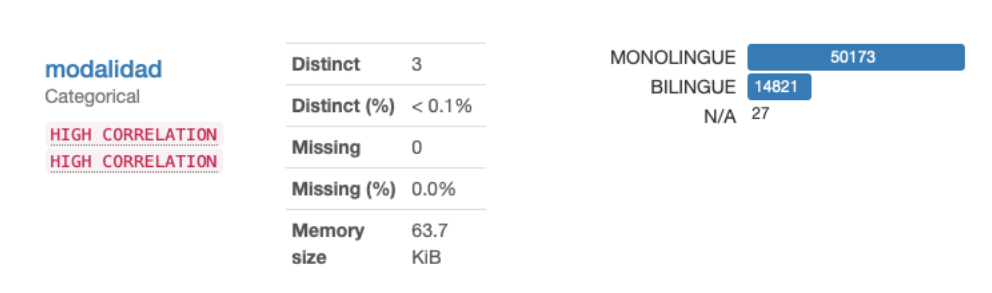
■ 6 variables distintas.



**Variable 12.** (*modalidad*) Es la modalidad en la que reciben las clases, si se imparten las clases en dos idiomas o solo uno.

**Valores posibles.**

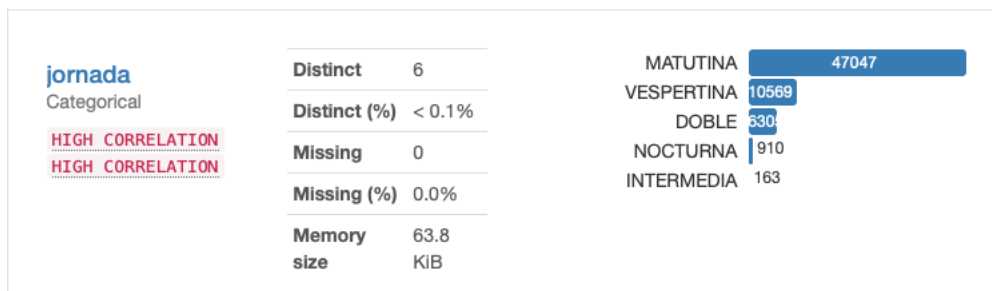
- 3 valores posibles.



**Variable 13.** (*jornada*) Es una variable que complementa a la variable de establecimiento ya que separa en una columna específica la jornada del centro educativo.

**Valores posibles.**

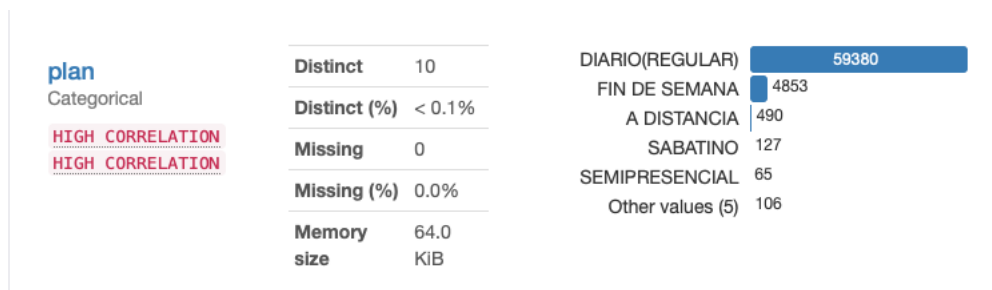
- 6 datos distintos.



**Variable 14.** (*plan*) Es una variable que hace referencia al plan en el que se reciben las clases de los centros educativos.

**Valores posibles.**

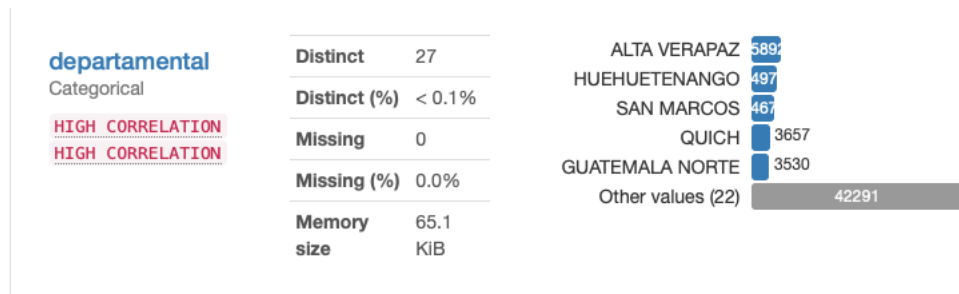
- 10 datos distintos.



**Variable 15.** (*departamental*) Es una variable que complementa a los distritos, ya que separa los centros educativos en varias subdivisiones; no necesariamente por departamento geográficos.

**Valores posibles.**

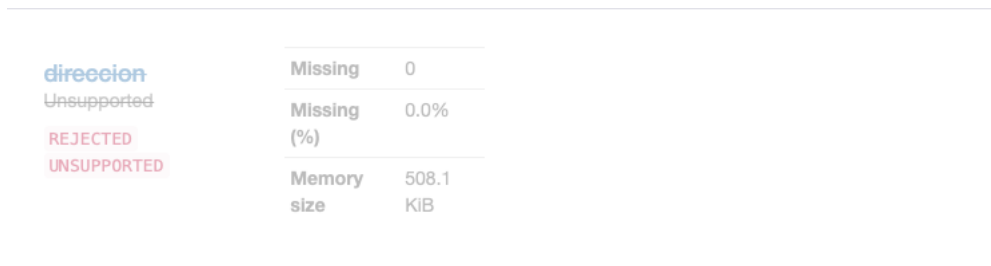
- 27 valores distintos.



**Variable 16. (direccion)** La dirección es una variable que contiene la dirección exacta de los centros educativos.

**Valores posibles.**

- Indeterminado, no parece ser una variable determinando en la base de datos.



**Variable 17. (telefono)** El número de teléfono es una variable que representa un medio de comunicación directa al centro educativo.

**Valores posibles.**

- Indeterminado, no parece ser una variable determinante en la base de datos. Esto se produce ya que se intentó tratar como variable categórica.

